



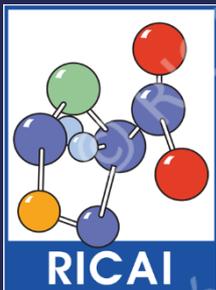
UNIVERSITÉ
DE LORRAINE

DHREAS



DÉTECTION DES INFECTIONS DU SITE OPÉRATOIRE PAR DATAMINING

Apport des données médicales issues du PMSI sur la performance
d'un modèle de détection



12 décembre 2016

Julie LIZON PharmD, MSc, Nicolas Jay, MD, PhD, Boris Marchard,
Arnaud FLORENTIN, MD, MSc

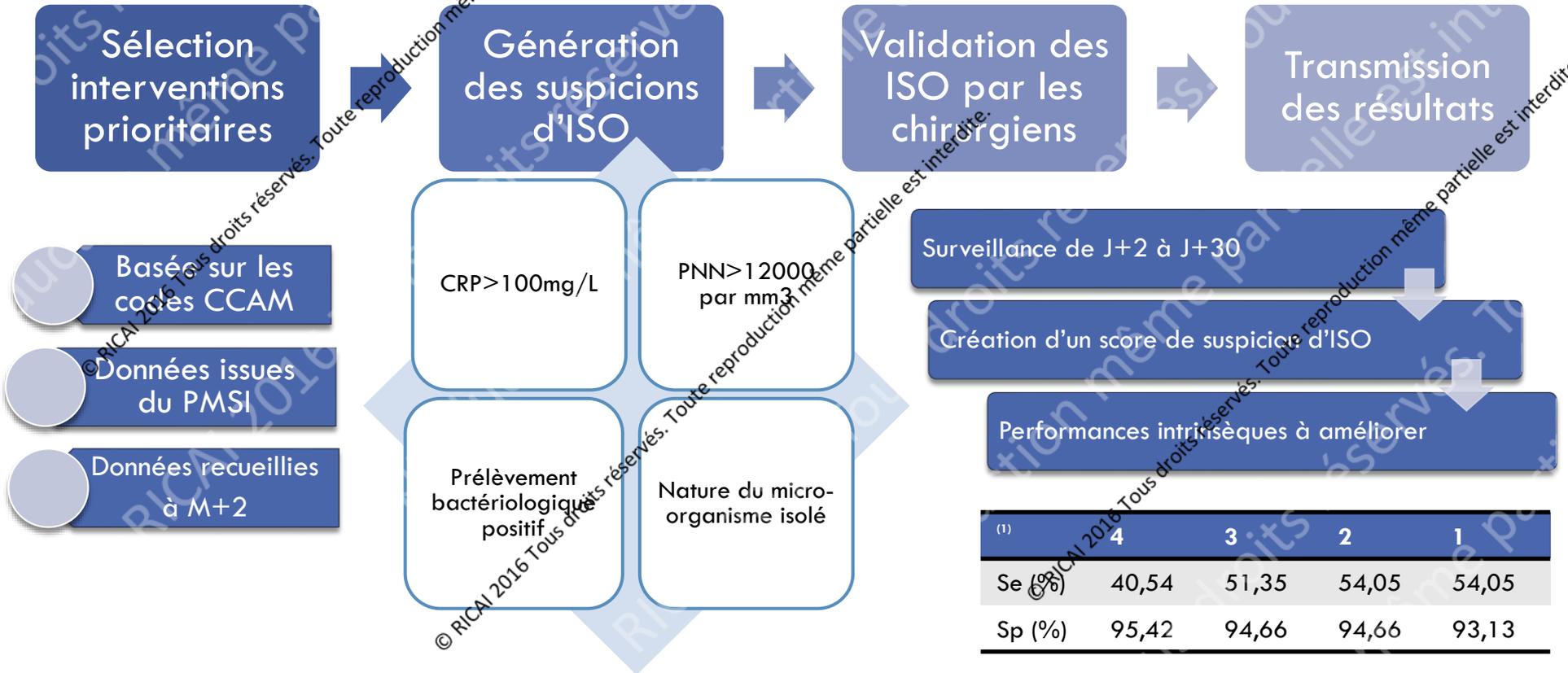
- Infection du site opératoire (ISO)
 - Infection survenant dans les **30 jours suivant l'intervention, ou dans l'année si il y a eu mise en place d'un implant, d'une prothèse ou d'un matériel prothétique**
 - **3^{ème} infections nosocomiales (IN) en France⁽¹⁾ avec une incidence de moins de 1% à plus de 15% selon type de chirurgie⁽²⁾**
 - **Conséquences** : séquelles, allongement durée hospitalisation, coût important, décès
 - **Prévention** : maîtrise du risque infectieux, surveillance des ISO

(1) Institut national de Veille Sanitaire (InVS). Enquête nationale de prévalence des infections nosocomiales et des traitements anti-infectieux en établissements de santé, France. 2012.

(2) Institut national de Veille Sanitaire (InVS). Surveillance des infections du site opératoire France, 2012. Résultats. 2012.

Organisation au CHRU de Nancy de la surveillance des ISO

3



(1) Hautemanière A, Florentin A, Hunter PR, Bresler L, Hartemann P. Screening for surgical nosocomial infections by crossing databases. J Infect Public Health. 2013 Apr;6(2):89–97.

Surveillance locale des ISO

4

- Difficultés d'organisation
 - Génération de faux positifs
 - Absence recueil données médicales pertinentes
 - Absence de détection de certaines ISO
 - Sensibilité médiocre
- Changement approche
 - Utilisation du datamining
 - Méthode exploratoire novatrice
 - Traitement grand nombre de données

Objectif principal

5

- Evaluer la contribution
 - ▣ Des données médicales issues du PMSI*
 - Diagnostics principaux (DP), Diagnostics Associées Significatifs (DAS), Codes CCAM
 - ▣ Et des modèles de prédiction de datamining pour détecter les ISO

Matériel et méthode

6

Population

Tous les adultes bénéficiant d'une intervention ciblée
entre le 1^{er} janvier et le 31 août 2015
10 services de chirurgie
Durée de suivi : 30 jours post-opératoire

Critère de jugement

ISO
Gold standard : diagnostic chirurgical

Sources des données

Données microbiologiques
Données biologiques
Données médico-administratives

Matériel et méthode

Datamining (DM)

Méthodes d'apprentissages supervisées

Modèle aléatoire

- Forêt aléatoire⁽¹⁾
 - Ajout d'une randomisation
 - Hasard dans choix des variables qui interviennent dans la construction des modèles

Modèle adaptatif

- AdaBoost⁽²⁾
 - Adaptation des modèles construits en donnant plus de poids aux observations mal ajustées ou mal prédites

(1) Breiman L. Random Forests. Mach Learn. 45(1):5–32

(2) Freund Y, Schapire RE. Experiments with a new boosting algorithm. Mach Learn. 1996;148–56.

Matériel et méthode

Evaluation des performances

Pour tous les modèles

- Se, Sp, VPP, VPN, Courbe ROC, AUC
- Indice NRI*(1)
- Calcul de l'erreur globale de classification

Pour les modèles DM

- Vérification pertinence construction modèle

* Net reclassification improvement

(1) Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008 Jan 30;27(2):157-72; discussion 207-12.

Résultats

9

□ Population

▣ 3900 patients inclus

▣ Moyenne âge : 56 ans +/- 19,79 [18-105]

▣ Sexe ratio : 0,78

	Nombre d'interventions	Détection ISO	ISO diagnostiquées	Taux d'ISO (%)
Chirurgie thoracique	118	24	16	13,54
Chirurgie gynécologique et obstétrique	639	4	3	0,47
Totalité	3900	405	192	4,92

Résultats

10

□ Performance modèles

	Algorithme actuel	Forêt aléatoire	AdaBoost
Sensibilité (%)	67,8	81,4	79,7
Spécificité (%)	92,2	97,8	97,7
VPP (%)	31,5	66,7	67,1
VPN (%)	98,1	99,0	98,9
AUC	0,827	0,988	0,979
 NRI	0	0,212	0,176
Erreur classification (%)	9	3	3

$p < 10^{-5}$

$p < 10^{-4}$ P=0,43

Discussion

- Construction modèles de détection
 - ▣ Injection nombreuses données médicales
 - 2497 variables explicatives
 - Sélection a priori diminue performance modèle de détection
 - ▣ Meilleures performances
 - Sensibilités, spécificités, VPP, VPN supérieures
 - AUC proche de 1
 - NRI positifs

Discussion

12

- Codage données
 - ▣ Identification codes intéressants
 - Evènement infectieux potentiel après la chirurgie
 - T814, L022, QZJA011
 - ▣ Amélioration codage
 - ▣ Evolution : text-mining
- En pratique
 - ▣ Evaluation échantillon indépendant
 - ▣ Automatisation

Merci pour votre attention

Surgical site infection detection by deaminating (SID₂): What is the contribution of medical data extract from hospital information system?

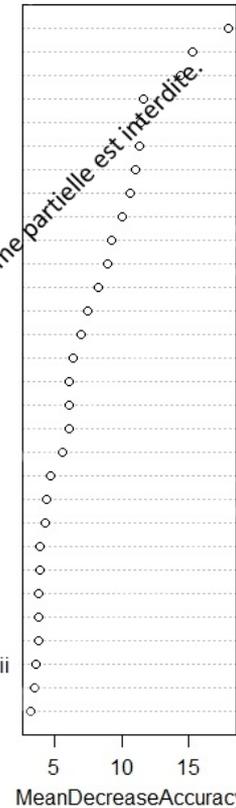
Julie Lizon PharmD^{1,2}, Nicolas Jay MD, PhD^{3,4}, Boris Marchand³, Arnaud Florentin MD, MSC^{1,2}

Soumis dans Journal of infection fin novembre 2016

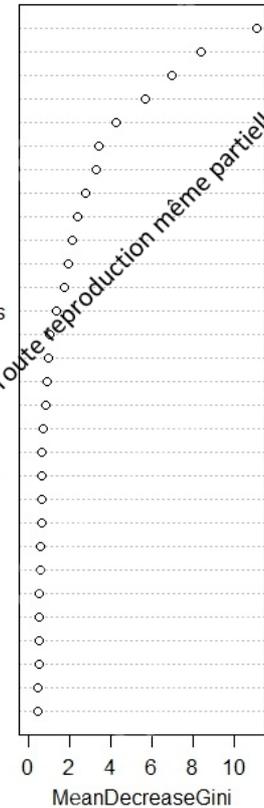
■ Pertinences variables

- T814 (Complication infectieuse d'actes de diagnostics ou thérapeutiques)
- L022 (Présence d'un abcès cutané)
- T8138 (Désunion de la plaie opératoire)
- QZJA011 (Evacuation collection profonde)

T814
L022
score_global
score_germes
Staphylococcus.aureus
score_plvt
Escherichia.coli
score_crp
Enterococcus.faecalis
T8138
Staphylococcus.epidermidis
Klebsiella.pneumoniae
Pseudomonas.aeruginosa
duree_hospit.5j
QZJA011
Morganella.morganii
L088
T81
N823
J188
score_pnn
HMQE005
ZCJA004
Clostridium.perfringens
J985
HZHE002
X0QK023
Corynebacterium.kroppenstedtii
Y834
var_coelio



T814
score_global
score_germes
score_plvt
score_crp
L022
T81
score_pnn
T8138
Escherichia.coli
Staphylococcus.aureus
Enterococcus.faecalis
Staphylococcus.epidermidis
K650
Y832
Klebsiella.pneumoniae
R65
B952
duree_hospit.5j
Pseudomonas.aeruginosa
U82
QZJA011
Y836
ZBQK002
T813
T8138
JZQP001
J18
L088
B957

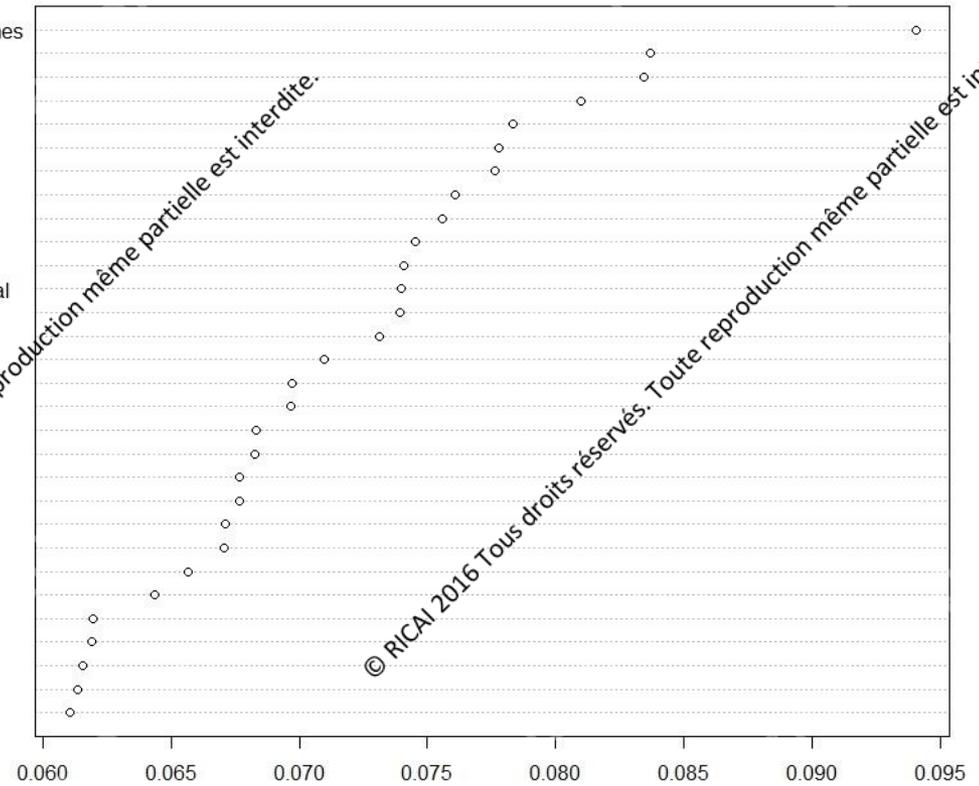


Pertinence des variables dans la constructions du modèle Forêt aléatoire

■ Pertinences variables

- L022 (Présence d'un abcès cutané)
- T8138 (Désunion de la plaie opératoire)
- T818 (Autre complication d'actes de diagnostics ou thérapeutiques)

score_germes
 score_pvt
 score_crp
 score_pnn
 Z590
 JJQX004
 HGQX008
 N61
 L022
 Y834
 T818
 score_global
 G25
 K413
 T8138
 H55
 G119
 AZQ005
 ZZQR002
 Z951
 T810
 N40
 C17
 Z902
 N823
 K409
 Z488
 Z8670
 EAQJ001
 Y836



Pertinence des variables dans la constructions du modèle AdaBoost